

Requested Patent:

JP7222248

Title:

SYSTEM FOR UTILIZING SPEECH INFORMATION FOR PORTABLE INFORMATION TERMINAL

10/17/00  
09/690313  
JC825 U.S. PTO



Abstracted Patent: JP7222248

Publication Date: 1995-08-18

Inventor(s): NENE YOSHITO; others: 01

Applicant(s): HITACHI LTD

Application Number: JP19940014214 19940208

Priority Number(s):

IPC Classification:

H04Q7/38 ; G06F3/16 ; G06F13/00 ; G06F15/02 ; G06F17/21 ; G10L3/00 ;  
G10L9/00

Equivalents:

**ABSTRACT:**

**PURPOSE:** To utilize a speech recognizing function at a portable information terminal by transmitting speech information through a communicating means to a server and returning the result of speech recognition to the portable information terminal as character information.

**CONSTITUTION:** At a portable information terminal 1, a speech is inputted by a speech input means 13, encoded and recorded. The recorded speech information is sent to a server 2 by a communicating means 15. At the server 2, the transmitted speech information is recognized by a speech recognizing means, converted to the character information and returned to the portable information terminal 1. Since processing for speech recognition is performed by the server 2 in sufficiently large hardware scale, a high-level speech recognizing method with a lot of calculation processing amounts can be utilized regardless of the hardware scale of the portable information terminal 1. When the speech information previously recorded in the portable information terminal 1 as a speech memory is used as the speech information to be transmitted to the server 2, a function for automatically converting the recorded speech information to characters can be provided.

(19)日本国特許庁 (JP)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開平7-222248

(43)公開日 平成7年(1995)8月18日

(51)Int.Cl. <sup>6</sup>	識別記号	序内整理番号	F 1	技術表示箇所
H 04 Q 7/38				
G 06 F 3/16	3 2 0 H	7323-5B		
13/00	3 5 4 D	7368-5B		
		7605-5K	H 04 B 7/26	1 0 9 H
		9288-5L	G 06 F 15/20	5 6 8 Z
審査請求 未請求 請求項の数 7 OL (全 7 頁) 最終頁に続く				

(21)出願番号 特願平6-14214

(71)出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(22)出願日 平成6年(1994)2月8日

(72)発明者 楠嶺 義人

東京都国分寺市東恋ヶ窪1丁目280番地

株式会社日立製作所中央研究所内

(72)発明者 天野 明雄

東京都国分寺市東恋ヶ窪1丁目280番地

株式会社日立製作所中央研究所内

(74)代理人 弁理士 小川 勝男

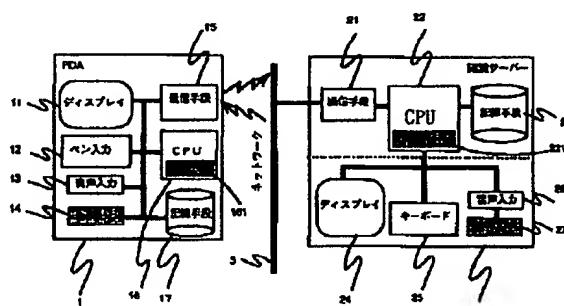
(54)【発明の名称】 携帯型情報端末における音声情報の利用方式

(57)【要約】

【構成】携帯型情報端末1に音声入力手段13と通信手段15を設け、ネットワークを介して通信可能な音声認識サーバ2上に、音声認識処理および認識結果の文字情報処理する手段を設ける。

【効果】サーバの処理に即時性が要求されない場合には、認識結果や文字情報処理結果を繰り返し利用して認識結果の正確さを向上させる方式などの、多くの処理時間を必要とする処理も利用可能になる。

図1



1

## 【特許請求の範囲】

【請求項1】音声入力手段と通信手段を有する携帯型情報端末において、入力された音声情報を前記携帯型情報端末内に記憶すると共に、その音声情報を通信機能を用いて別のサーバー計算機に転送し、前記サーバー計算機上において前記音声情報を入力とする音声認識処理と、前記音声認識処理の結果に対する文字情報処理を行い、その結果得られた文字情報を、前記携帯型情報端末に送り返して、前記携帯型情報端末上に前記文字情報を表示することを特徴とする、携帯型情報端末における音声情報の利用方式。

【請求項2】請求項1において、音声情報を転送する直前に、前記音声情報に対して施したい音声認識処理および文字情報処理の内容を、リクエスト情報として指定あるいは選択する機能を有する携帯型情報端末における音声情報の利用方式。

【請求項3】請求項1において、前記音声認識処理は、処理に必要なパラメータを複数個用意し、前記各パラメータを用いた複数の処理を実行し、その結果得られた複数の音声認識結果を比較して、その中の一つを認識結果として出力する音声認識処理である携帯型情報端末における音声情報の利用方式。

【請求項4】請求項3において、前記パラメータは、音声信号中の音声区間を切り出すための時間窓の長さである携帯型情報端末における音声情報の利用方式。

【請求項5】請求項1において、前記音声認識処理には、単語単位で認識結果を確定する処理を用い、かつ認識結果の単語を確定する際に、あらかじめ設けておいたしきい値よりも低い確率で確定した単語を未確定単語とし、前記未確定単語の認識結果候補を全て記憶しておき、一度全ての入力音声に対する処理が終了した後に、前記未確定単語よりも後に出現した認識結果の情報を用いて、前記未確定単語の確定を行う携帯型情報端末における音声情報の利用方式。

【請求項6】音声入力手段と通信手段を有する携帯型情報端末で利用する情報管理ソフトに於いて、入力したい情報を音声で入力し、入力された音声情報を前記携帯型情報端末内に記憶し、前記音声情報に対応する絵記号をディスプレイ上に表示し、前記音声情報を通信機能を用いて別のサーバー計算機に転送し、前記サーバー計算機上で前記音声情報を入力とする音声認識処理と、キーワード抽出処理と、前記キーワード抽出処理の結果に対する文字情報をスケジュール表の表示に適合する順番に並び変える処理を行い、その結果得られた文字情報を、前記携帯型情報端末に送り返して、前記携帯型情報端末上に表示された情報管理表の上に前記文字情報を表示する携帯型情報端末における音声情報の利用方式。

【請求項7】請求項6に於いて、入力音声情報に対応する絵記号を表示するディスプレイ上の位置情報、およびサーバー計算機より送り返される文字情報を表示するディ

10

20

30

40

50

2

スプレイ上の位置情報は、音声を入力する時に、同時にペン入力手段を用いて前記ディスプレイ上の位置を指示することにより得られる携帯型情報端末における音声情報の利用方式。

## 【発明の詳細な説明】

## 【0001】

【産業上の利用分野】本発明は携帯型情報端末装置における音声情報の利用方式に係り、特に、通信機能を介して実現される音声認識機能の利用方式に関する。

## 【0002】

【従来の技術】近年、個人用の情報機器の一つとして、携帯型情報端末装置の開発が盛んに行われている。

【0003】この携帯型情報端末装置は、従来の卓上型パーソナルコンピュータとは異なり、キーボードを備えていないのが一般的である。文字情報の入力は、液晶画面上にペンなどを利用して手書きで行う。コマンドは液晶画面上に表示されるアイコンを選択すること等により行われる。また携帯型情報端末には、有線または無線による通信機能が設けられている。この通信機能は、電子メールによる文字情報の送受信や、FAX機能による図形情報の送受信に用いられる。さらに、音声を録音する機能を持つ携帯型端末もある。録音された音声は、音声メモとして保持され、必要に応じて再生できる。

【0004】ところで、このような携帯型端末では、キーボードがないため、文字情報やコマンドの入力にペン入力を用いているが、音声認識機能を搭載し、音声によるコマンド入力などが行える方が、使い勝手が向上することが予想される。このため、将来の携帯端末に搭載する機能として、音声認識機能を挙げる将来像の例は数多く提案されている。

【0005】実際に、卓上型パーソナルコンピュータでは、音声認識機能を搭載したものが製品化されており、いくつかのコマンドを不特定話者の音声で実行することができる。この技術を用いれば音声認識機能を携帯型情報端末に搭載することは可能である。しかし音声認識機能を搭載した携帯型情報端末の例は現時点ではない。

【0006】現時点の音声認識技術レベルは、認識を行う単位を単語とし、認識結果はあらかじめ辞書に登録されている単語の中で、もっとも確からしいものが選択されるという程度に留まっている。また、雑音の混入や話者の違いなどにより、認識率が大きく低下する。連続に自由に発声した音声を完全に文字化することは、現時点の技術では不可能である。このような要求に答えるには、音声認識技術の一層の高度化が必要になる。

## 【0007】

【発明が解決しようとする課題】従来の携帯型情報端末では、音声の録音機能は搭載されているものの、記録された音声情報はそのまま再生する音声メモとして利用されるだけであり、録音された音声を自動的に文字として表示する利用方式については、考慮されていなかった。

【0008】一方、一般に音声認識機能の高度化に伴い、それに必要な計算処理量は増加する。しかし携帯型情報端末では、使用できるハードウェアに限界があるため、認識可能な単語数を制限するなど、実現できる音声認識機能のすべてを搭載できないという問題があった。

【0009】本発明の目的は、ハードウェア規模の限られる携帯型情報端末で、録音された音声情報を自動的に文字化するための手段と、大規模なハードウェアを必要とする高度な音声認識機能を利用するための手段とを提供することにある。

【0010】

【課題を解決するための手段】上記の課題は、携帯型情報端末が通信できるネットワーク上に、音声認識手段を有する大規模なハードウェアをもつ情報処理装置（以下サーバと略す）を設け、携帯型情報端末から入力した音声情報を、通信手段によりサーバに送り、サーバ内で音声認識を行い、認識結果を文字情報として表現し、携帯型情報端末に送り返すことにより解決できる。

【0011】

【作用】携帯型情報端末では音声を入力し符号化して記録する。記録された音声情報は通信機能によりサーバに送られる。サーバでは音声認識手段により、送られてきた音声情報を認識し文字情報に変換する。音声認識のための処理は、ネットワーク上のいずれかに設置され、ハードウェア規模が十分に大きいサーバで行われるので、携帯型情報端末のハードウェア規模に関係なく、計算処理量が多い高度な音声認識手法を利用することが可能となる。

【0012】また、サーバに送る音声情報として、あらかじめ携帯型情報端末内に、音声メモとして記録されていた音声情報を用いれば、録音された音声情報を自動的に文字化する機能を実現することができる。

【0013】

【実施例】図1は本発明のシステムのブロック図である。携帯型情報端末（以下PDAと略す）1は、演算処理装置（以下CPUと略す）16の周辺に、液晶ディスプレイなどの表示手段11、ペン入力手段12、音声入力手段13、音声出力手段14、情報記憶手段17、およびサーバが接続されているネットワークに無線又は有線により接続可能な通信手段15を有する。

【0014】一方、音声認識サーバ2は、CPU22の周辺に、ネットワークに接続するための通信手段21、および大容量の情報記憶手段23とを有する。この他にサーバには、管理者がサーバを管理するために必要な、ディスプレイなどの表示手段24、キーボードなどの情報入力手段25、音声入力手段26、および音声出力手段27が備えられている。

【0015】本実施例では、PDA上に実現される機能、およびサーバ上に実現される機能は、各々の内部に設けられたCPUで実行されるソフトウェア161およ

び221で実現されている。もちろん一部の演算を専門に行う手段をCPUとは別に設け、ソフトウェアの実行速度を高速化する方式もある。

【0016】PDA内のCPU上で実行されるソフトウェア161は、一般のパーソナルコンピュータと同様、階層的な構造を持ち、基本ソフトまたはオペレーティングシステム（以下OSと略す）と呼ばれるものと、そのOSの機能を利用して種々の機能を実現するアプリケーションソフトがある。また、このOSとアプリケーションとの間には、ウインドウ環境制御ソフトがある。ユーザからはアプリケーションソフトの機能しか見えないが、アプリケーションソフト内の機能を実現するために、ウインドウ環境ソフトやOSの持つ種々の機能が働いているという点は、一般のパーソナルコンピュータと同じである。

【0017】本発明のPDAの特徴は、音声入力手段により入力された音声情報を、情報記録手段に記録すると同時に、記録された音声情報に対応する図形（以下アイコンと表わす）などを表示手段に視覚的に表示し、さらにユーザからの要求があった場合には、この記録された音声情報を通信手段を通じて音声認識サーバへ転送する、という機能を実現するソフトウェアを有する点にある。もちろん記憶された音声情報は、ユーザからの要求があった場合には、そのまま音声として再生する機能もある。

【0018】図2に本システムを使用するために必要な、PDAおよびサーバ上で動作させるソフトウェアの手順をフローチャートとして示す。

【0019】PDA上のソフトウェアは、ウインドウ環境ソフトの一部として、あるいは独立したアプリケーションソフトとして提供される。本ソフトウェアには、送信する音声情報に対して、ユーザが施したい処理内容を指定する機能があり、ここで指定された処理はリクエスト情報として、音声情報のデータに付加される。ここで指定する処理とは、キーワード抽出、フォーマッティング、機械翻訳などがあり、たとえば、メニューとしてディスプレイ上に表示された中から、ペンによって選択する。また、音声認識サーバに音声認識処理の方式が複数個ある場合、どの方式を実行するかを指定する機能もある。

【0020】一方、本実施例の音声認識サーバは、一般的なサーバ用ワークステーションなどによって構成される。したがって、基本的なハードウェアおよびソフトウェアの構成には新規な点はない。本発明のサーバの特徴は、ネットワーク等を介した通信手段によって入力された音声情報を分析し、その音声情報に含まれる言語情報を抽出するための音声認識手段を有する点にある。

【0021】音声認識手段の実現方式には種々の方式が考えられるが、本発明のサーバに用いられる音声認識手段の場合には、入力音声情報を繰り返し利用する方式を

用いることができる。すなわち、通信手段を介してPDAより送られてきた音声情報を認識する場合には、実時間処理を行う必要がないこともある。この場合には認識結果を決定するのに十分な時間を使うことができる。

【0022】具体的には、図3に示したように、入力音声部分を切り出すための信号処理パラメータを変えながら、複数回の認識処理を行い、その結果得られた複数の認識結果の内、最も確からしいものを最終的な認識結果として出力する方式がある。

【0023】また、一度入力音声情報全体の認識結果を決定した後、確定度が低く一つの結果に決定しきれなかった部分、例えば、単語を抽出し、その部分の前後の高い確定度で決定した部分の情報をを利用して、未確定部分を確定する方式もある。

【0024】図4はこの処理の概念を示す図である。同図は最初の認識処理の結果確定できなかった部分(図中では?の列で示す)を、一度全ての入力音声に対する認識処理が終了した後に、未確定単語よりも後ろにある確定部分の類似単語を用いて確定する例を示している。さらに、複数の雑音低減処理手段を用意し、入力音声情報に対して、雑音低減処理手段を変えて複数回認識処理を行い、その中で最も確定度が高かったものを、最終的な認識結果として出力する方式もある。

【0025】また本発明のサーバは、音声認識手段によって得られた言語情報から、キーワードを検出する手段も有している。本手段は、携帯端末から送られてきた音声情報に付加されているリクエスト情報に要求があった場合に動作し、キーワードを抽出し出力する。抽出するキーワードの種類、例えば、地名や日付などの指定もリクエスト情報によって制御できる。本手段の出力をPDAに表示することにより、入力された音声情報を簡略化して文字表示することが可能となる。

【0026】さらに本発明のサーバは、音声認識手段によって得られた文字情報を、他の言語に翻訳する機械翻訳手段を有している。本手段は、携帯端末から送られてきた音声情報に付加されているリクエスト情報に要求があった場合に動作し、例えば、音声認識結果の文章を英語の文章に変換する。

【0027】この機械翻訳手段に加えて、音声合成手段をサーバに設け、音声情報をサーバ側からPDA側に送ることができるようにすれば、PDAより入力した音声を文字化し、他の言語に翻訳し、さらに音声化することで、音声翻訳機能あるいは通訳機能を実現することが可能になる。サーバでの各処理に時間がかかる場合には、実時間の通訳機能は実現できないが、あらかじめ用意したスピーチなどを他国語に変換して保存しておくなどの使い方ができる。

【0028】図5には、本発明のシステムの利用方式の一例を示す。PDA1のユーザ4は、音声でスケジュールを入力する。PDA1は、内部の情報記憶手段、例え

ばメモリーにこの音声情報を記録し、表示手段上11に入力した音声情報に対応するアイコン111を表示する。ここでこの音声情報を文字化し、キーワードを抽出するようユーザが指示を付加した場合には、音声情報データは上記の要求を指示するリクエスト情報が付加され、通信手段を介して音声認識サーバ2に転送される。音声認識サーバでは、音声認識手段により音声情報を文字化し、リクエスト情報にしたがって、キーワード抽出手段によりキーワードを抽出して、PDAに転送する。PDAでは、送られてきた文字情報112を表示手段上の適切な位置に表示する。これによりユーザは、音声で入力したスケジュールが、入力された時点では音声情報としてしか参照することができなかったが、一定の時間後には、キーワードを文字として表示手段上で参照することができる。

【0029】さらに、音声入力と同時にペン入力手段を用いて、ディスプレイ上の位置情報を入力すれば、これら2種類の入力情報を統合的に利用することが可能となる。たとえばスケジュール入力の際、ディスプレイ上にカレンダーを表示し、音声アイコンおよびサーバでの音声認識結果の文字情報を表示させたい位置を、ペンによって指定する、などの利用方式が考えられる。

#### 【0030】

【発明の効果】本発明によれば、PDA上で入力した音声情報を、音声として参照できると同時に、文字としても参照することができる。

【0031】また、PDAのハードウェア規模では実現が不可能な音声情報に対する大規模な処理を、PDAで利用することができる。

【0032】また、サーバの処理に即時性が要求されない場合には、認識結果や文字情報処理結果を繰り返し利用して認識結果の正確さを向上させる方式などの、多くの処理時間を必要とする処理も利用可能になる。

【0033】さらに、従来よりPDAに備えられているFAX機能を利用すれば、文字化された音声情報を、携帯型端末よりFAXとして送ることができるので、結果的に音声をFAXで送ることができる。

#### 【図面の簡単な説明】

【図1】本発明のシステムのブロック図。  
【図2】PDA側で行う音声情報処理のフローチャート。

【図3】入力音声情報を繰り返し利用する音声認識処理の説明図。

【図4】認識結果を繰り返し利用して認識結果を修正する処理の説明図。

【図5】本発明の利用方式の一例を示す説明図。

#### 【符号の説明】

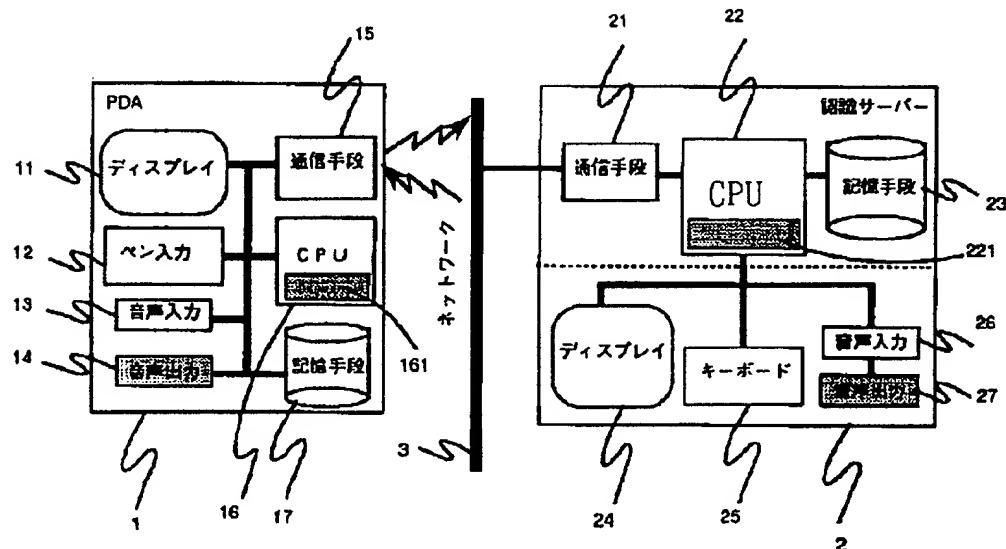
1…携帯型情報端末、11…情報表示手段、12…文字情報入力手段、13…音声入力手段、14…音声出力手段、15…通信手段、16…演算処理装置、17…情報

記憶手段、161…プログラム、2…音声認識サーバ、  
21…通信手段、22…演算処理装置、23…情報記憶  
手段、24…情報表示手段、25…文字情報入力手段、

26…音声入力手段、27…音声出力手段、221…ブ  
ログラム、3…ネットワーク設備。

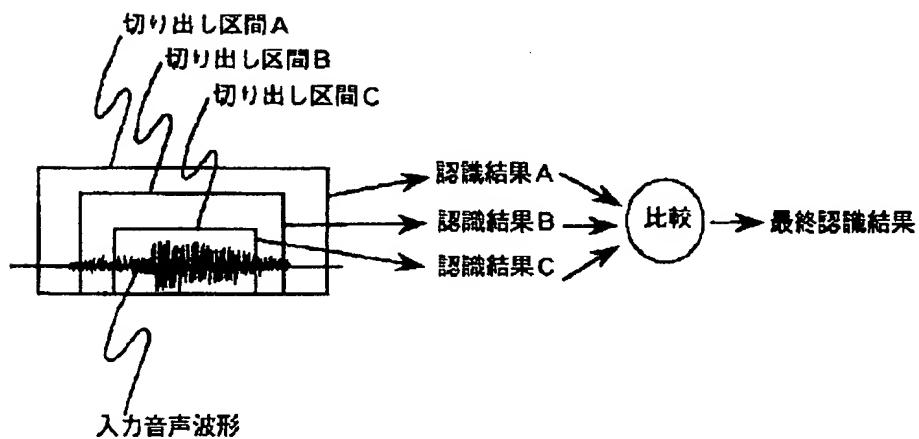
【図1】

図1



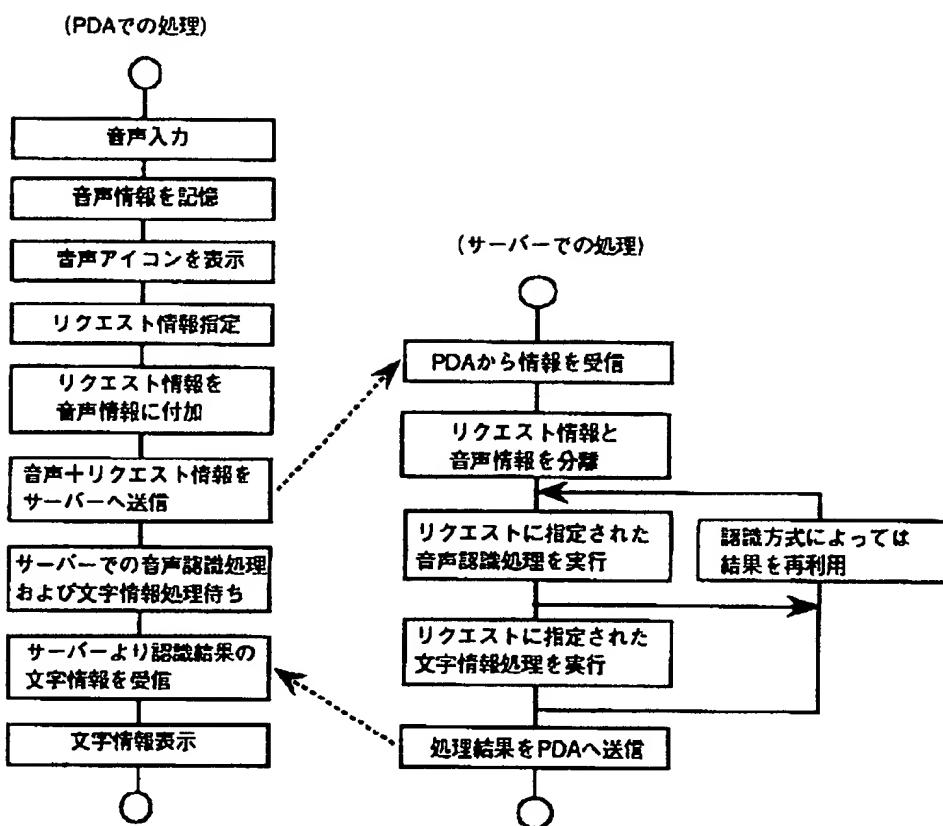
【図3】

図3



[図2]

図2



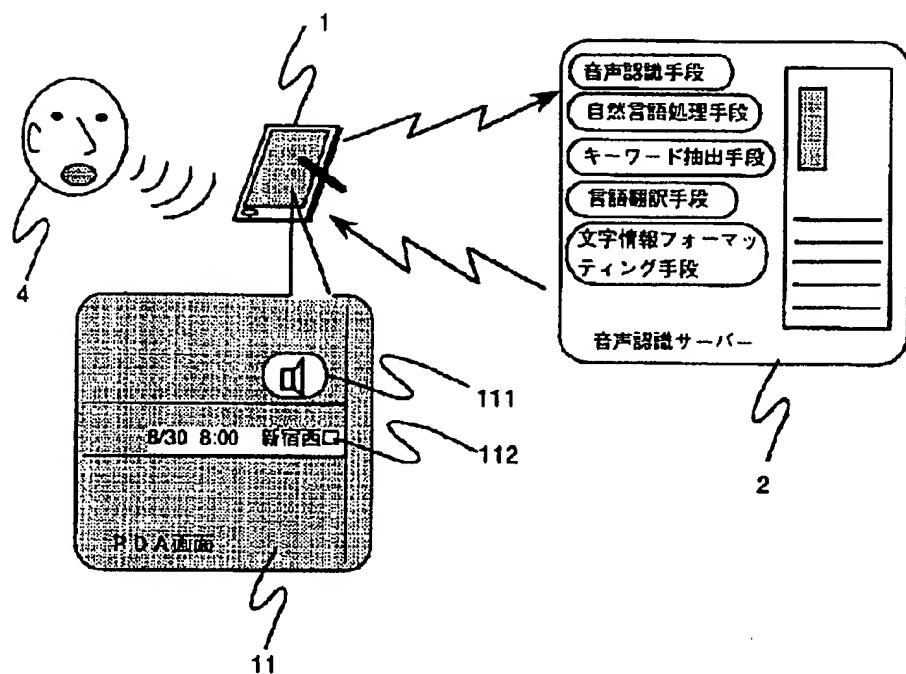
[図4]

図4



【図5】

図5



フロントページの続き

(51) Int.Cl. <sup>6</sup>	識別記号	序内整理番号	F I	技術表示箇所
G 06 F 15/02	3 2 5 B			
	3 3 5 E			
17/21				
G 10 L 3/00	5 3 1 C			C6
	5 5 1 A			C1
	5 6 1 C			C5
	E			C5
	5 7 1 H			C2
	G			C7
9/00	3 0 1 A			C4
		7605-5K	H 04 B 7/26	1 0 9 M